



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

How stable are personalities? A multivariate view of behavioural variation over long and short timescales in the sheephead swordtail, *Xiphophorus birchmanni*

Citation for published version:

Boulton, K, Grimmer, AJ, Rosenthal, GG, Walling, CA & Wilson, AJ 2014, 'How stable are personalities? A multivariate view of behavioural variation over long and short timescales in the sheephead swordtail, *Xiphophorus birchmanni*', *Behavioral Ecology and Sociobiology*, vol. 68, no. 5, pp. 791-803.
<https://doi.org/10.1007/s00265-014-1692-0>

Digital Object Identifier (DOI):

[10.1007/s00265-014-1692-0](https://doi.org/10.1007/s00265-014-1692-0)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Behavioral Ecology and Sociobiology

Publisher Rights Statement:

This is a PDF file of an unedited manuscript that has been accepted for publication. The publisher version is available at: [10.1007/s00265-014-1692-0](https://doi.org/10.1007/s00265-014-1692-0)

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



How stable are personalities? A multivariate view of behavioural variation over long and short timescales in the sheephead swordtail, *Xiphophorus birchmanni*.

Kay Boulton¹, Andrew J Grimmer², Gil G Rosenthal³, Craig A Walling^{1*}, Alastair J Wilson^{2*}

¹ Institute of Evolutionary Biology, University of Edinburgh, West Mains Rd, Edinburgh, EH9 3JT, UK

² Centre for Ecology and Conservation, Biosciences, College of Life and Environmental Sciences, University of Exeter, Cornwall Campus, Treliever Road, Penryn, Cornwall, TR10 9EZ, UK

³ Department of Biology, Texas A&M University, 3258 TAMU, College Station, TX 77843, USA and Centro de Investigaciones Científicas de las Huastecas “Aguazarca”, Calnali, Hidalgo, Mexico.

*Authors made equal contributions

Key words: multivariate behaviour, personality, repeatability, stability, boldness, I matrix

Contact: Alastair J Wilson: A.Wilson@exeter.ac.uk

Please cite this article as: Boulton, K, Grimmer, AJ, Rosenthal, GG, Walling, CA & Wilson, AJ, 'How stable are personalities? A multivariate view of behavioural variation over long and short timescales in the sheephead swordtail, *Xiphophorus birchmanni*' Behavioral Ecology and Sociobiology (2014), <http://dx.doi.org/10.1007/s00265-014-1692-0>

Abstract

Many studies have revealed repeatable (among-individual) variance in behavioural traits consistent with variation in animal personality; however, these studies are often conducted using data collected over single sampling periods, most commonly with short time intervals between observations. Consequently, it is not clear whether population-level patterns of behavioural variation are stable across longer timescales and/or multiple sampling periods, or whether individuals maintain consistent ranking of behaviours (and/or personality) over their lifetimes. Here we address these questions in a captive bred population of a tropical freshwater poeciliid fish, *Xiphophorus birchmanni*. Using a multivariate approach, we estimate the among-individual variance-covariance matrix (**I**), for a set of behavioural traits repeatedly assayed in two different experimental contexts (open field trials, emergence and exploration trials) over long- (56 days between observations) and short-term (four day observation interval) time periods. In both long- and short-term data sets we find that traits are repeatable and the correlation structure of **I** is consistent with a latent axis of variation in boldness. While there are some qualitative differences in the way individual traits contribute to boldness, and a tendency towards higher repeatabilities in the short term study, overall we find that population-level patterns of among-individual behavioural (co)variance to be broadly similar over both time frames. At the individual level we find evidence that short-term studies can be informative for an individual's behavioural phenotype over longer (e.g. lifetime) periods. However statistical support is somewhat mixed and, at least for some observed behaviours, relative rankings of individual performance change significantly between data sets.

39 Introduction

40 It is now apparent that, within animal populations, individuals often exhibit differences in behaviour
41 that are repeatable across time and context. This repeatable variation is taken as evidence for
42 animal temperament (e.g. Boissy 1995; Réale et al. 2007), behavioural syndromes (Sih et al. 2004),
43 coping styles (Koolhaas et al. 1999), or personality, the latter term reflecting parallels with research
44 in human psychology (Budaev 1997b; Gosling 2001). A number of axes of among-individual
45 behavioural variation condensed into “personality traits” have been described, including boldness-
46 shyness, exploration-avoidance and general activity (Réale et al. 2007). Understanding the evolution
47 of personality has become a major field of study in behavioural ecology (Dall et al. 2004; Stamps and
48 Groothuis 2010). There is now growing evidence that traits relating to personality contribute to
49 fitness variation and therefore may be both adaptive and generally under selection (Smith and
50 Blumstein 2008). However, if natural selection occurs through variation in lifetime fitness, then an
51 important question arises: just how stable are personalities over individual lifetimes? Here we
52 address this question in a captive population of fish. We do this using a novel multivariate approach
53 that characterises personality variation as a latent character underpinning among-individual
54 (co)variation in a suite of observed behaviours.

55 While there remains considerable disagreement over how best to define individual personality traits
56 (Réale et al. 2007; Toms et al. 2010; Carter et al. 2013; see below) there is broad consensus that
57 among-individual behavioural variance is the statistical signature of animal personality. Typically this
58 is quantified as the (among-individual) repeatability, defined as the proportion of observed variance
59 explained by individual identity, of one or more observed behavioural traits. Thus partitioning of
60 observed variance into among- and within-individual components (the latter arising from individual
61 plasticity and/or measurement errors) from repeat observations on individuals is crucial to empirical
62 studies of personality (Dingemanse et al. 2012b; Brommer 2013; Araya-Ajoy and Dingemanse 2014).
63 In a meta-analysis, Bell et al (2009) concluded that on average, estimates of repeatability for

observed behavioural traits decreased as the interval between sampling events increased. Consequently, it may be dangerous to assume that short-term studies reflect behavioural (and by implication, personality) differences that are stable over the lifetime of individuals. This is potentially important since short-term repeatability estimates predominate in the literature, although the number of studies conducted over timeframes that may be considered more representative of natural life-spans is growing (for more recent examples, see Ronning et al. 2005; Bushuev et al. 2010; Chervet et al. 2011; David et al. 2012; Kanda et al. 2012). However, few studies have collected repeated observations over two distinct time periods from the same individual (but see for e.g. Carere et al. 2005) that would allow the “repeatability of repeatability” to be assessed. Here we do this, but also extend our analysis to the multivariate case to ask whether patterns of among-individual behavioural (co)variation reflect an underlying personality trait that is stable across distinct long- and short-term sampling periods.

In what follows we investigate the temporal stability of multiple behavioural traits in the freshwater poeciliid fish, *Xiphophorus birchmanni* to answer two complementary questions. Firstly, at the level of the population, how stable are the patterns of among-individual trait (co)variance generated by underlying personality? Secondly, at the level of the individual, do short term studies reveal behavioural tendencies that are stable across lifetimes? To answer these questions we characterise behavioural variation along what we loosely consider to be an axis of shyness-boldness. Boldness is the most commonly studied axis of personality in fish (Toms *et al.* 2010), and positively correlates with fitness-related traits including reproductive success, parental provisioning, growth, aggression, social dominance, dispersal and proactive responses to stressors such as predation risk (Dingemanse et al. 2004; Brown et al. 2005; Bell and Sih 2007; Cote et al. 2010; Rudin and Briffa 2011; Ariyomo and Watt 2012; Mutzel et al. 2013). There remains, however, a lack of consensus on how best to define boldness and how it should be assayed (Toms et al. 2010). This raises obvious potential for misclassification of personality traits (Carter et al. 2013), and/or disagreement over appropriate experimental design (Toms et al. 2010).

The present goal is to investigate stability of a personality trait without adding further to existing debate over issues of definition. Consequently we do not attempt to define boldness or the best way to measure it *a priori*; rather, we follow the view of others that personality traits should be considered as latent variables that can best be uncovered by observing several measurable, correlated and potentially overlapping behaviours across contexts (Dochtermann and Jenkins 2007; Dingemanse et al. 2010; Dochtermann and Roff 2010). We therefore make a distinction throughout between “behavioural traits” that are observed directly, and “personality (traits)”, inferred from among-individual (co)variance in observed behaviour(s). This exploratory approach, which follows Huntingford (1976) and others (Budaev 1997b; Moretz 2003) is becoming more mainstream and allows the avoidance of difficulties that can arise if a single behaviour is chosen *a priori* to assay boldness. For example, a fish that swims a long distance in one behavioural trial may be classified as willing to explore and therefore as “bold”; however, this behaviour could also plausibly be indicative of anxiety, with the animal’s “exploration” being driven by a search for refuge.

Currently the most common experimental paradigm used to measure boldness is that of the open field trial (OFT), in which an animal is placed in an open arena and its behaviour is monitored for a predetermined observation period. Initially developed for rodent studies (Hall 1934; Walsh and Cummins 1976), OFTs have long been applied to fish models (Warren and Callaghan 1975; Budaev 1997b). Considered the most reliable way to assay boldness by some authors (Burns 2008), others have argued that OFTs risk conflating boldness with other axes of variation that are distinct (if sometimes correlated) personality traits in their own right (e.g. exploration-avoidance, overall activity, Réale et al. 2007). If so, then simple modifications to OFTs such as providing a refuge from which an animal can choose to emerge and explore (emergence and exploration trial, EET) may be useful (Dingemanse et al. 2007).

In what follows we use both types of behavioural trial mentioned above (OFT and EET) to observe how fish behave in these contexts and to characterise the repeatable component of multivariate

behaviour. We then assess the extent to which one or more major axes of variance adequately depict observed variation. In other words, we aim to describe the behavioural trait variation first, and then consider the extent to which its repeatable component fits within the paradigm of a major axis of personality, i.e. the boldness-shyness axis (Dingemanse et al. 2010; Dochtermann and Roff 2010). We then go on to address three specific questions regarding the temporal stability of personality. Firstly we ask whether repeatabilities estimated from repeated measures of individual behaviours over a short time period give a misleading view of the importance of among-individual variance over longer time periods. Secondly, by extending our analysis to the multivariate case we ask whether the structure of the between-trait among-individual covariance matrix, denoted \mathbf{I} , following Wilson et al. (2013), is similar when estimated from short- and long-term data; i.e. do repeated empirical analyses of a single population actually reveal the same major axes of among-individual variation? If so, then a final question concerns the extent that individuals retain the same relative ranking for repeatable behaviours, and hence personality, over their lifetimes.

Methods

Study species and husbandry

One hundred wild adult *Xiphophorus birchmanni* were caught in the Arroyo Coacuilco near the town of Coacuilco, municipality of San Felipe Orizatlán, Hidalgo, Mexico, (elevation 314 m lat/long 21.099 -98.587), and imported to the UK in February 2010. Between August 2010 and May 2011 we collected an offspring generation ($n = 384$) from 13 males and 27 females (mean (SE) brood size of 8.86 (0.541)). Gravid females were isolated and, following birth, broods were immediately netted and moved to one half of a partitioned 30 L tank; broods of more than six offspring were split with each half of the family placed in different tanks. Fry were fed twice daily on a mix comprising equal quantities of crushed ZM spirulina and brine shrimp flake and laboratory prepared brine shrimp

nauplii. At an average of 17 weeks (range 12 to 27) juveniles were tagged with a single elastomer injection for individual identification purposes and transferred to mixed-family rearing groups of $n = 8$. Note it is not possible to determine sex at this age in this species and therefore the sex ratio was not controlled. Eight rearing groups were then kept within each of six sequentially set-up stacks of tanks, each stack sharing a common water supply and recirculating filtration system. As part of a parallel study of density effects on growth, rearing groups were initially housed under two different density regimes as follows. Within each stack, four groups were placed in 30 litre (37 x 37 x 22 cm) glass tanks (low density treatment) with the remaining four groups in 15 L half tanks (high density treatment). Half tanks were created by placing a black net covered Perspex-framed partition down the centre of a full – size tank. Thus, establishing a stack required 64 fish (i.e., 8 x 8) to be available for tagging simultaneously and this accounts for the variation in tagging age within stacks. Fish were fed twice daily with a standardised ration of flake food as above (morning) and a mix of previously frozen blood worm and daphnia (afternoon). On the days when behavioural data was to be collected, the morning feed was omitted in an attempt to encourage exploration tendencies. Temperature was maintained between 22 - 24°C and a 12:12 hr light:dark cycle imposed. After being housed in this manner for 28 weeks, density was swapped for half of the tanks, thus creating four treatment effects with the total number of fish divided approximately equally between them as follows: Low/Low ($n = 93$), Low/High ($n = 95$), High/High ($n = 87$), High/Low ($n = 93$). Observations from individuals failing to reach sexual maturity by the end of the long-term study (50 weeks), were excluded from the analysis and the above breakdown ($n = 11$).

Behavioural data collection

The trials were performed over two experimental study periods, denoted long-term (LT) and short-term (ST). All available fish contribute to the long-term data set ($n = 373$) while a random subset of 32 fish from each of the four density treatments (Low/Low $n = 13$, Low/High $n = 4$, High/High $n = 9$, High/Low $n = 6$) was used for the short-term study (Table 1). Trials were of two types, open field

(OFT) and emergence and exploration (EET) with multiple specific behavioural traits assayed in each trial type (Table 2). Overall, the long-term trials took 13 months to complete (May 2011 – May 2012), with data collected over an actual 30 week period for each fish. Each individual was subject to an OFT followed by an EET seven days later, a process that was repeated three times at 56 day intervals, thus yielding four OFT and four EET trials per fish. The short-term data set was collected in February 2013, with 32 individual fish subjected to alternating OFT and EET at 48 hour intervals (i.e. 2 days between trials, 4 days between repeated trials of the same type) with each animal undergoing five trials of each type. For those 32 individuals used in both study periods data was therefore collected over a timeframe with a mean (SE) of 531.4 (6.38) days. By comparison the mean (SE) longevity of individuals with known birth and death dates under our laboratory conditions is 450.3 (8.10).

Experimental procedures

Open Field Trial (OFT)

An empty 45 x 25 x 25 cm tank was filled to a depth of 8 cm with room temperature water (22°C). The tank was lit from below and visually screened by a cardboard casing to prevent external laboratory disturbance. Fish were caught individually from their home tank with a dip net, quickly examined for identification tags and immediately placed into the centre of the OFT tank. Following a 30 second acclimation period, behaviour was filmed for 300 seconds using a Sunkwang C160 video camera fitted with a 5 – 50 mm manual focus lens suspended above the tank. Data were then extracted from the video using the tracking software Viewer II (<http://www.biobserve.com/products/viewer/index.html>), which was set up to divide the tank basal area into two approximately equal halves (middle and perimeter zones) (Fig. 1a). Water was changed between individual trials to prevent chemical cues affecting behaviour.

Emergence and Exploration Trial (EET)

A 45 x 25 x 25 cm tank was physically divided into three sections with opaque Perspex, providing a right-hand, centre and left-hand chamber. A small (5 cm) opening was cut in each divider, starting two cm from the tank edge. The openings were positioned at opposite sides of the tank. The chamber on the right hand side was designated as the refuge, and equipped with a plastic plant and several small stones. A rising trapdoor was rigged to a pulley above the tank, and positioned inside the refuge and covering the exit into zone 1 (Fig. 1b). Tanks were filled, emptied, lit and screened as above. Fish were individually caught and examined as before, and placed directly into the centre of the refuge where they were allowed 30 seconds to acclimate before the trapdoor was lifted. Filming then commenced for 300 seconds (as above), but only behaviour outside the refuge (i.e. in zones 1 and 2) was tracked and extracted for analysis.

Behavioural traits

The behavioural traits recorded in this study were selected as those likely to reflect variation along a bold-shy type personality axis. For the OFT, we predicted that fish tending toward boldness would actively explore the novel environment of the OFT by leaving the tank sides and spending more time in the central zone than shy fish. OFT behaviour was therefore quantified by four traits; Track Length (TL), Activity (Act), Area Covered (AC) and Time in Middle of the tank (TIM), which we predicted would be positively correlated with one another. In the EET, we expected bold fish to locate the doorway in the refuge and leave through it. We recorded two traits from the EET: whether or not the individual emerged from the relative safety of the refuge (Emergence) and Latency in seconds to do so. We predicted positive within-individual correlations between Emergence from the refuge and the OFT traits, with negative correlations between Latency to Emerge and all other traits. Note that the EET tank was set up with the area outside the refuge further divided into two zones (1 and 2; Fig 1b). In the EET, we had initially planned to use “latency

to enter zone 2" (distal to the refuge) as an additional trait in our analyses; however, in practice this became a redundant trait due to a low frequency of fish entering this area.

Statistical analyses

All data were modelled using restricted maximum likelihood mixed effects models implemented in ASReml V3 (Gilmour et al. 2009). Prior to analysis, data for the OFT trait Time in Middle were square root transformed to reduce positive skew. Visual inspection of residuals suggested that the assumption of residual normality was reasonable for the other traits recorded in OFT. All traits were rescaled to standard deviation units prior to analysis to prevent trait scale effects from influencing the structure of \mathbf{I} (defined and estimated as described below). Given that a large proportion of fish did not emerge from the refuge (see results) the Latency to Emerge data were heavily censored and we elected to use only the binary variable of Emergence in subsequent analyses. Emergence was included in full multivariate models using REML under an assumption of (multivariate) residual normality. Statistical inferences on this trait should therefore be treated with obvious caution.

While statistical approaches exist that allow non-Gaussian trait distributions to be used (e.g. MCMC Bayesian approaches implemented in the R package MCMCglmm (Hadfield, 2010)) they do not currently allow the error structures appropriate to our multivariate models (i.e. no definable or estimable residual covariance between OFT and EET traits – see below) and thus could not be used here. However, we checked the validity of REML-based conclusions regarding Emergence by fitting additional univariate and bivariate models using MCMCglmm. Specifically we fitted a univariate model of Emergence to estimate the repeatability of this trait and bivariate models of Emergence with all other OFT traits to estimate the covariance structure between these traits. All models in MCMCglmm modelled Emergence as a categorical trait with the residual variance fixed at 1 and all OFT traits as Gaussian. All MCMCglmm models were run for a total of 1050000 iterations with a burnin of 50000 iterations and a thinning interval of 1000 iterations. The repeatability of Emergence from MCMCglmm models was defined as the intraclass correlation, calculated as $V_I / (V_I + V_R + \pi^{2/3})$,

where V_I is the among-individual variance and V_R is the residual variance that in this case is fixed to 1 (Hadfield 2010).

To test the hypothesis that among-individual variance for behavioural traits is both present and repeatable in our fish species, we first combined data from both collection periods and fitted a multivariate model of our observed behavioural traits. For each trait we included fixed effects of the *mean*, *sex* (a two level factor determined from external morphology at maturation), *home stack* (a six level factor accounting for differences between sets of fish sharing water supplies), *trial number*, *density treatment*, and *day order*. *Trial number* is the cumulative number of trials experienced by an individual (fitted as a linear effect). *Density treatment* is a four level factor describing density conditions experienced in the rearing stacks. *Day order* was modelled as a linear effect of the number of preceding trials conducted on any day and was used as a proxy for time of day. This was included to control for potential diurnal rhythms in fish behaviour. We also fitted an interaction term of *trial number* * *density treatment*, in case any systematic changes in observed trait means across trials (due to e.g., age effects, habituation etc.) are themselves treatment dependent. Wald F-tests were used to test the significance of fixed effects in the models.

By including individual identity as a random effect, we then partitioned multivariate phenotypic (co)variance not explained by the fixed effects into an among-individual and a within-individual (residual) component. The former is estimated as the variance-covariance matrix **I**, which contains estimates of the among-individual variance (V_I) component for each trait on the diagonal and estimates of the corresponding covariance between trait pairs (COV_I) off the diagonal. The within-individual component is similarly estimated as a residual variance-covariance matrix (**R**). We make the standard assumptions that residual errors are normally distributed and uncorrelated across observations, and that (co)variance parameters in **I** and **R** are homogeneous across levels of the fixed effects (i.e. density treatments, trial number, stack etc). Although the two experiment-specific sets of traits are not observed in the same trials, we grouped the data by trial period, (e.g. OFT1 with

EET1). Thus, we modelled a residual covariance term between OFT and EET traits observed within each trial period. Repeatability (R_i) was then estimated for each trait as the among-individual variance (V_i) divided by total phenotypic variance (V_p) (where V_p is the phenotypic variance conditional on the fixed effects; i.e. $V_p = V_i + V_R$). Between each pair of traits (1, 2) the among-individual covariance (COV_i) was rescaled to give the corresponding correlation r_i (where $r_{i(1,2)} = COV_{i(1,2)} / \sqrt{(V_{i1} * V_{i2})}$).

To test the statistical significance of among-individual behavioural variation we compared the likelihood of our full multivariate model to two further models. In the first of these, we fitted I as a diagonal matrix such that the model allows among-individual variance V_i for each trait, but assumes Cov_i is zero between all trait pairs. In the second, a null model, we removed the random effect of individual identity completely. Comparison of the diagonal model with the null model using likelihood ratio tests (LRT) allows a global test of the significance of among individual behavioural variance (Wilson et al. 2010). Comparison of the full model with the diagonal model, again by LRT, allows a statistical test of whether I contains significant between-trait covariance structure (Wilson et al. 2013). LRT were performed by estimating χ^2_{nDF} as twice the difference in model log likelihoods, with the number of degrees of freedom (n) equal to the number of additional parameters to be estimated in the more complex model.

The above analyses were then repeated using long- and short-term data subsets to estimate the corresponding matrices I_{LT} and I_{ST} and associated parameters. Note that, following the conclusion of the LT, the density treatments were no longer applied and the 32 fish used in the ST were housed together in the same stack. Therefore, the fixed effect *stack* was redundant and omitted from the models for the short-term subset analyses. To further investigate the structure of I , I_{LT} and I_{ST} , we subjected each matrix to eigenvector (EV) decomposition. This allowed us to examine: a) how much variance is captured by the first axis (EV1) of multivariate behaviour in each case, b) whether factor loadings of individual traits onto EV1 are consistent with an interpretation of “boldness-shyness”

and c) whether EV1 is similar in \mathbf{I}_{LT} and \mathbf{I}_{ST} . To provide a quantitative measure of how similar the multivariate behavioural axes emerging from the long- and short-term data sets were, we calculated the angle (θ) between the first eigenvectors of \mathbf{I}_{LT} and \mathbf{I}_{ST} . An angle of $\theta = 0^\circ$ equates to the vectors being perfectly aligned, meaning that EV1, i.e. the axes of multivariate behavioural variation in \mathbf{I}_{LT} and \mathbf{I}_{ST} are identical. Conversely, an angle of $\theta = 90^\circ$ would indicate the vectors are orthogonal (and thus maximally differentiated) to each other across the two different time periods (i.e. the major axis of behavioural variation across the two studies are independent).

Uncertainty around the factor loadings for individual traits on EV1 (for \mathbf{I} matrix) and around θ was estimated using a parametric bootstrap approach (similar to that outlined in the appendix of (Morrissey et al. 2012)). We simulated 5000 replicate draws of \mathbf{I} , \mathbf{I}_{LT} and \mathbf{I}_{ST} from multivariate normal distributions using the maximum likelihood estimates of these matrices as the means, and the variance-covariance matrices of their elements to define the variances. In each case the 5000 simulated matrices were subject to Eigen decomposition. Uncertainty around the point estimates of trait-specific factor loadings was then described using the 95% highest probability density interval the simulated values of these loadings (for \mathbf{I} , \mathbf{I}_{LT} and \mathbf{I}_{ST} respectively). Note that these intervals should be viewed as approximate as they are vulnerable to departures from multivariate normal assumptions. By comparing 5000 pairs of simulated LT and ST matrices we similarly estimated the uncertainty around our point estimate of θ . Note however that since θ cannot be less than zero, we also generated a null distribution for the estimator in the absence of any difference between (true) \mathbf{I} matrices. This was done by comparing the leading eigenvector of each of the 500 replicate draws of \mathbf{I}_{LT} (simulated as described above), to the leading eigenvector of a second matrix, simulated with the same mean (i.e. the REML point estimate of \mathbf{I}_{LT}) but a variance equal to the estimated variance-covariance matrix from the short-term study. Thus the null distribution represents θ estimates given that i) the angle is zero since true \mathbf{I} matrices are identical (and equal to the REML estimate of \mathbf{I}_{LT} , but ii) the second (short-term) matrix (and so its leading eigenvector) is estimated with greater uncertainty due to the lower sample size.

Finally, we compared V_I estimates in LT and ST data subsets, and tested the among-individual, across data subset correlations ($r_{I(LT,ST)}$). For each behavioural trait (x) we used a likelihood ratio test to compare a bivariate model of x_{LT} and x_{ST} where V_I is constrained to be equal, to a model where it is free to vary. This tests the hypothesis that among individual variance differs across data sets. (Note that since traits are analysed in observed standard deviation units V_I can also be interpreted as the repeatability estimate unconditional on fixed effects). We then expanded this model to estimate the among-individual, across data subset correlation ($r_{I(LT,ST)}$) and tested this against null hypotheses of both $r_I = 0$ and $r_I = +1$. Estimation of this correlation is possible since the 32 fish used in the short-term study were a subset of the long-term study. If $r_I = +1$, then this indicates that the ranking of phenotypic merits (i.e. each individual's repeatable component of the observed trait) is the same across data sets. However, if $r_I = 0$, then an individual phenotypic merit in the long-term study is uncorrelated with the repeatable component of that same behaviour observed over a short time period in later life.

Results

In total, 1235 sets of behavioural observations were conducted from a possible 1492, the difference being due to mortality of some fish over the study period. Summary data for all behavioural traits are presented in supplemental materials, Fig. S1. In EET, the number of fish emerging from the refuge within the observation period was lower than anticipated based on pilot data (LT = 526/2448, ST = 100/318), resulting in severe censoring of Latency to Emerge data. We therefore elected to use only the binary Emergence trait from this trial type in our analyses.

Analysis of full data set

There was significant among-individual variance in multivariate behaviour (diagonal model versus null model, $\chi^2_5 = 125.6$, $P < 0.001$), as well as among-individual covariance among traits (diagonal

model versus full model, $\chi^2_{10} = 101.8$, $P < 0.001$). Estimates of individual repeatability ($R_i (\pm SE)$) were low to moderate, ranging across traits from 0.055 (± 0.024) for Emergence (on the observed scale, estimated by REML) to 0.192 (± 0.029) for Time in Middle (Table 3). Based on univariate models, V_i was statistically significant at $P < 0.05$ for all traits (Supplemental Table S2). The estimated fixed effects are not directly relevant to present objectives; however they are presented in full in the supplemental materials (Supplemental Table S3).

Between traits, the signs of all among-individual correlations (r_i) were positive, consistent with our *a priori* expectations (Table 3). The OFT traits Track Length, Activity and Area Covered were all strongly correlated (and nominally significant based on $|r_i| > \text{two standard errors}$); however while Time in Middle was strongly correlated with Area Covered ($r_i = 0.653 \pm 0.075$, Table 3), it was only weakly associated with the other OFT traits. The EET trait Emergence was positively correlated with each OFT trait (r_i estimates ranging from 0.304 with Track Length to 0.577 with Activity, Table 3).

Eigen analysis of I , estimated from the full data set revealed that the first two vectors explained 64 % (eigenvector 1, EV1) and 26 % (eigenvector 2, EV2) of the repeatable among-individual variation respectively (Fig. 2). The trait loadings on the dominant vector EV1 are consistent with an interpretation of this axis of variation as boldness (or arguably exploration and/or general activity; see discussion). Thus individuals that tended to emerge repeatedly in the EET, swim longer distances, are more active explore more area, and spend more time in the middle of the OFT tank. By comparison, EV2 trait loadings show this axis to be dominated by time spent in the middle of the tank. Track Length and Activity load on this vector to a lesser extent and with an opposing sign to Time in Middle, while the other traits show limited contributions to EV2 (Fig. 2b).

As noted earlier, our REML analysis makes an assumption of (multivariate) residual normality that is violated by inclusion of the binary trait Emergence. Univariate analysis of Emergence using MCMCglmm, calculated following equation 15 of (Nakagawa and Schielzeth 2010), yielded a slightly higher estimate of repeatability (on the liability scale) with a posterior mode of $R = 0.090$, 95% HPD

interval 0.024 – 0.177, Table S1). While noting that interval will never span zero since R is constrained to lie in positive parameter space, the posterior mode is nonetheless “distinct” from zero (Supplemental material, Fig. S2). Bivariate models (i.e. the use of one OFT trait plus Emergence as the phenotypic variates) also confirmed the presence of strong positive among-individual correlations (r_i) between Emergence and OFT traits. Thus, the MCMCglmm analyses corroborate the results of the REML analysis for Emergence (Supplemental material, Table S1).

Comparison of long- and short-term results.

In both the long- and short-term studies, the presence of repeatable variance was statistically supported (comparisons of null and diagonal model: LT $\chi^2_5 = 77.0$, $P < 0.001$; ST $\chi^2_5 = 29.7$, $P < 0.001$) as was the presence of between-trait among-individual covariance structure (comparisons of diagonal and full multivariate model: LT $\chi^2_{10} = 95.0$, $P < 0.001$; ST $\chi^2_{10} = 54.9$, $P < 0.001$). Univariate models confirmed that V_i was statistically significant for all OFT traits in both LT and ST, but not for Emergence in ST (Supplementary Tables S2).

The estimate of I_{LT} is very similar to that obtained using all data (as described above), not unexpected given that the long-term study contributes the bulk of the total data set. However, comparison of I_{LT} and I_{ST} (and derived parameters thereof) indicates some differences in the structure of among-individual behavioural variation as estimated from our long- and short-term studies (Table 3). Note that the smaller size of the short-term data set means that the estimates are less precise for this study; this is reflected in the larger standard errors associated with the parameters. Repeatability estimates (R) were higher in the short term study across all traits. However the increased R from ST was particularly striking for Track Length (Table 3, Fig. 3). For this trait, along with Activity and Area covered the null hypothesis of equality of (V_i) across data sets could be rejected (comparison of bivariate models with homogeneous and heterogeneous V_i , $P < 0.05$, Fig. 3).

The among-individual between-trait correlations (r_i) reveal a broadly similar structure for the long- and short-term studies (Table 3). Thus estimates for ST largely confirm our *a priori* expectation of positive correlation structure between the OFT traits and Emergence. One qualitative exception to the expected pattern is provided by Time in Middle. In LT this trait is positively correlated with all other traits as expected; however, in ST the sign of r_i is negative (but not significant) between Time in Middle and Track Length and Emergence (Table 3).

Eigen decomposition confirms the view that qualitative differences between I_{LT} and I_{ST} are largely related to Time in Middle. Thus, in both data sets the first eigenvector again dominates the variance in I (accounting for 66% and 73% in long- and short-term respectively), consistent with an important latent character underlying behavioural variation (Fig. 2a). Time in Middle has a strong positive loading on $EV1_{LT}$, consistent with our *a priori* expectation that a bold fish would spend more time in the middle of the open field arena, the corresponding loading coefficient is close to zero (in fact slightly negative) on $EV1_{ST}$. The angle (θ) between $EV1_{LT}$ and $EV1_{ST}$ is 34.63° (95% HPD interval, 5.03- 53.09 $^\circ$). While the point estimate of 34.63° indicates at least some divergence between the leading eigenvectors on a scale from 0 (no difference) to 90 (axes are orthogonal), it is not significantly greater than the angle expected by chance if the true matrices are identical (95% HPD of the null distribution for θ generated by our parametric bootstrap is from 1.54 – 69.14 $^\circ$). While we acknowledge that our null distribution indicates low statistical power to reject the null hypothesis that $\theta = 0$ (see Supplemental Fig. S3), our conclusion is however that $EV1_{LT}$ and $EV1_{ST}$ are broadly similar, with qualitative differences largely attributable to the decreased loading of TIM on $EV1_{ST}$. This is further evidenced by a drop in θ from 34.63° to just 11.15° for the corresponding comparison of I estimates excluding Time in Middle. There are also some qualitative inconsistencies evident between $EV2_{LT}$ and $EV2_{ST}$ for the OFT traits, due to greater loadings on Track Length (changes sign), Activity, Area Covered and Time in Middle, while the loading on Emergence is reduced (also changes sign) (Fig. 2b). The angle (θ) between $EV2_{LT}$ and $EV2_{ST}$ = 48.32° (95% HPD interval 25.75-86.48 $^\circ$), which again is not significantly different from null expectations.

For those individuals tested in both long- and short-term studies, the among-individual correlations between LT and ST data sets were positive (although not always significant based on likelihood ratio tests) for OFT traits (Fig. 4), ranging from 0.219 (\pm 0.294) to 0.729 (\pm 0.314). Estimates were significantly greater than zero for Area Covered and Time in Middle. However, we also found that the correlation was significantly less than 1 for the traits Track Length and Activity. Thus, while phenotypic performance of an individual in one data set may be predictive of its behaviour in the other, there is also evidence that the ranking of individuals, at least for Track Length and Activity, significantly differs between long and short term studies. For Emergence the corresponding among-individual correlation estimates between long- and short-term were actually negative, though not significantly so. In fact the estimate was characterised by so much uncertainty that despite being negative it was not possible to reject the null hypothesis of $r = +1$. We suggest this is a result of the low repeatable variation of Emergence and thus little weight should be placed on this result.

Discussion

Data from our long-(LT) and short-term (ST) studies provide evidence of among-individual variance in behaviour, both when considered separately and in combination. Of the five traits assayed in the two distinct types of behavioural trial - open field (OFT) and emergence and exploration (EET) - repeatabilities were statistically supported in all cases. In addition our analyses support the presence of a significant among-individual correlation structure for behavioural traits in I. Correlation structure is found both within- and across-contexts (i.e. trial types), indicating behavioural variation among fish that is consistent with accepted definitions of animal personality. We found that repeatabilities of OFT traits were higher than the EET though not significantly so in all cases. Our results therefore support the assertion of Burns (2008) that the OFT is a good and reliable test of boldness and exploratory behaviour in small fish, although it is certainly possible that the EET could be better optimized to target the among-individual component. We discuss the

biological interpretation of (multivariate) variance within these two trial types further below. However, here we note the pragmatic consideration that the binary distribution of Emergence data obtained from the EET is more difficult to analyse and interpret while the censoring of Latency to Emerge created a data distribution not readily modelled in any software. Although such problems are likely surmountable by modification of the behavioural assay (e.g., using an extended observation time to eliminate or at least reduce censoring), at least in this case it is not clear to us that the EET provides additional biological insight.

Comparison of long- and short-term data sets suggested that the patterns of individual (co)variance between traits frequently used to define boldness are relatively stable. Nevertheless, as predicted *a priori* we found a tendency for the magnitude of R_i to decrease with a higher interval between observations, at least in OFT trials. For example, repeatabilities for OFT traits ranged from 0.188 to 0.458 in the short term data (with repeat observations at an average interval of four days) but 0.136 to 0.207 in the long term data (average interval of 56 days). In a meta-analysis of behavioural repeatability studies that included either long- (i.e. > 1 year) or short-term (i.e. < 1 year) intervals between observations, the average (median) across all estimates was 0.37 (Bell et al. 2009). Here our repeatability estimates pertain to correlated traits and are therefore not independent. Nevertheless, apart from our short-term study estimates for Track Length and Activity, we note that our estimates for all other traits were lower than those of the meta-analysis average. Repeatability estimates from short-term studies in the meta-analysis (Bell et al. 2009) outnumbered those from long term studies by 11:1; however, our study considers observations collected within two distinctly separate periods across individual lifetimes.

Arguably the more important question to be asked of our long- and short-term data sets concerns the stability of correlation structure within the multivariate \mathbf{I} matrix and the interpretation of boldness from its eigenvector decomposition. As seen with the single trait repeatabilities, the structure of \mathbf{I}_{LT} mirrored that of \mathbf{I} estimated from all data combined. This is unsurprising given that

the long-term data comprised a much greater number of individuals and will thus dictate patterns in the combined dataset. I_{LT} is dominated by a single vector that is broadly consistent with our expectations of boldness. Significant within- and between- trial type correlations indicate that individuals emerging from the EET refuge are more likely to have high scores for all OFT traits, thus matching our expectation of bold behaviour.

Though not statistically significant, qualitative differences between I_{LT} and I_{ST} were apparent. These differences were focussed around the sign and strength of correlations between Time in Middle and traits from both trial types, indicating that both bold and shy individuals from the short-term study spent a similar amount of time in the middle, whereas in the long-term study, shy fish had behaved in a more thigmotaxic manner. This pattern was reflected in comparisons of the major eigenvectors of long- and short-term data, where a moderate, albeit not statistically significant, angle (θ) between the first long- and short-term axes was estimated. Furthermore, if Time in Middle is dropped from the calculation, the estimated angle is reduced by more than half. Thus our interpretation is that both data sets reveal a major vector of among-individual (co)variance in observed behavioural traits. This vector is similar in the two data sets and can be interpreted as a latent personality trait - namely boldness. In both data sets bolder individuals tend to swim longer distances, be more active and explore more area (in the OFT), and are more likely to emerge from a refuge (in the EET). However, tendency to spend more time in the middle of the OFT arena appears not be a reliable indicator of boldness as it was only associated with this vector in the LT study. Indeed this trait was the major source of qualitative difference between the two matrices.

In the current study it is not possible to distinguish whether higher repeatabilities and the changing structure of I with regard to Time in Middle are a consequence of the sampling period (long- vs. short-term) or potentially reflect interesting, possibly even species-specific, biological changes that happen with age and/or trial experience. Note, however, that our analyses control for any habituation effects on mean behaviour, and that we found little statistical support for individual-by-

trial-number interactions (results not shown). More generally some authors have argued that individual behaviour is likely to become more rigid and follow more set patterns over time (Roberts and DelVecchio 2000). If so we would predict increasing repeatabilities with age (here confounded with time scale of data collection). Conversely, others suggest that in the absence of any disturbance (e.g., in a constant laboratory environment), expectations of changes to individual patterns of behaviour formed in early life are ill-founded (Stamps and Groothuis 2010). While no overall differences were found between juvenile and adult behavioural repeatabilities in the Bell et al. (2009) meta-analysis, a subset of data suggested juvenile behaviour to have higher repeatability than that of adults. However, the metanalysis contained only three studies that included observations following individuals through from juvenile to adult status. Thus direct comparison of age classes is not straightforward. Clearly more empirical studies of how repeatability changes with age would be valuable, as indeed would parallel studies exploring environmental dependence. Here we assumed homogeneous variance structures across environments (density treatments, stacks) and other fixed effects (sexes, day order) for simplicity. These assumptions can be relaxed in the statistical models to test for and quantify individual by environment (IxE) as changes in the among-individual variance (or structure of I in the multivariate case) (Dingemanse et al. 2010). Here *post hoc* analyses of the LT data set provides some evidence of heterogeneous repeatabilities across density treatment classes (see Supplemental Table S4). Though not expected to bias current conclusions (parameter estimates presented are effectively averaged across treatments), if robust this effect may certainly be biologically interesting.

The population level patterns of among-individual (co)variances between traits were broadly similar between LT_i and ST_i, albeit with some differences as described above. However, by using the same individuals in both long- and short-term studies we were able to address the question of whether the relative ranking of individuals with respect to their behavioural tendencies was stable. The estimates of r_i for each observed behavioural trait between the long- and short-term datasets provide a mixed answer to this question. Positive correlations for the OFT traits do show a degree of

stability in (repeatable) behavioural tendencies across the data sets though statistical support was mixed and it appears individuals were more likely to maintain a consistent ranking for some traits (e.g. Area Covered) than others (e.g. Track Length).

We previously stated it is not our intention to be prescriptive about what boldness is or how it should be assayed. Nevertheless, *a priori*, we anticipated that in the OFT, bold fish would travel long distances and be willing to visit a large area of the tank including the central zone, and that these traits would correlate significantly with whether individuals emerged in the EET. However, this depiction requires that the bold individual is also active and/or exploratory. Above we have noted that the major axis of variation in I is largely consistent with expectations of a bold-shy continuum as the terminology is used in the literature; however, the strength of among-individual correlations suggests that it could equally be called exploration or general activity in a novel environment. Nevertheless, as qualitatively almost all the variance loads onto this single axis of variance, we conclude that these continuums (personality axes) are, at least in our study species, either the same entity or so tightly correlated that attempting to distinguish between them may have little practical value. Indeed, Burns (2008) concluded that emergence from a refuge was difficult to interpret strictly as either boldness or exploration, even though it has been described as boldness only by others, (e.g. Budaev 1997a; Brown et al. 2005). Exploring the functional significance of the consequences of this behavioural variance in wild populations is likely to yield more insight than further debate with regard to terminology (e.g. Dingemanse et al. 2012a; Kurvers et al. 2012; Carvalho et al. 2013). Nonetheless, we have sufficient statistical support in our results to conclude that both trial types revealed behaviours characteristic of boldness, evident from the strong among-individual correlations between all the observed traits. This again leads us in the direction of Burns' (2008) view that in practice, the OFT offers the most useful test arena for this axis of personality. Here we have obtained repeated measures of multiple behavioural traits during two test types and across two distinct sampling periods (long- versus short-term), something that has seldom been accomplished in the literature. In practical terms, we conclude that the open field trial is preferable

to the emergence and exploration trial as an experimental test for investigating boldness, and we show how eigen decomposition of an I matrix can usefully identify latent personality traits. This multivariate approach is broadly similar to that used in several other recent studies (Budaev 2010; Carter et al. 2013; Araya-Ajoy and Dingemanse 2014). Our study also provides information about the stability of personality, both in terms of population level patterns and individual differences. We find that observed behavioural traits are repeatable over long time periods as well as when observations are made over only a few weeks, although there is a tendency for short term estimates to be higher. Taking a multivariate approach we show that I is dominated by a single vector through phenotypic space that is similar across the two study periods and can be interpreted as boldness. We note however, there are at least some qualitative differences in the relationships of observed behaviours to this vector. At the individual level we also find qualified support for the proposition that short-term studies are informative for an individual's behavioural phenotype over longer (e.g. lifetime) periods.

Ethical standards

Ethical review committees at the Universities of Edinburgh and Exeter approved all work in this study, which was carried out under license granted by the Home Office (UK) under the Animals (Scientific Procedures) Act 1986.

Acknowledgements.

The authors would like to thank Niels Dingemanse and three anonymous reviewers for helpful comments that have helped to improve this manuscript. Additionally we thank Jarrod Hadfield for advice with MCMCglmm software. This work was supported by an EPSRC studentship to KB and a BBSRC fellowship to AJW. CAW was supported by a NERC Junior Research Fellowship.

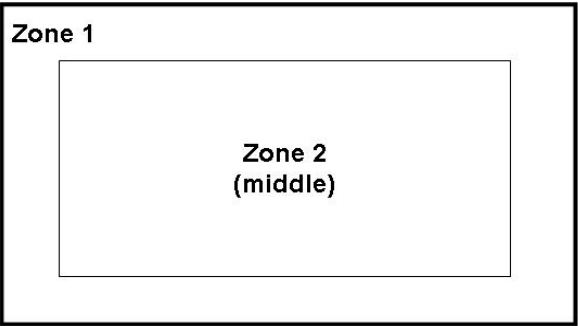
References

- Araya-Ajoy YG, Dingemanse NJ (2014) Characterizing behavioural 'characters': an evolutionary framework. *Proc R Soc B* 281:20132645
- Ariyomo TO, Watt PJ (2012) The effect of variation in boldness and aggressiveness on the reproductive success of zebrafish. *Anim Behav* 83:41-46
- Bell A, Sih A (2007) Exposure to predation generates personality in threespined sticklebacks (*Gasterosteus aculeatus*). *Ecol Lett* 10:828-834
- Bell AM, Hankison SJ, Laskowski KL (2009) The repeatability of behaviour: a meta-analysis. *Anim Behav* 77:771-783
- Boissy A (1995) Fear and fearfulness in animals. *Q Rev Biol* 70:165-191
- Brommer JE (2013) On between-individual and residual (co)variances in the study of animal personality: are you willing to take the "individual gambit"? *Behav Ecol Sociobiol* 67:1027-1032
- Brown C, Jones F, Braithwaite V (2005) In situ examination of boldness-shyness traits in the tropical poeciliid, *Brachyrhaphis episcopi*. *Anim Behav* 70:1003-1009
- Budaev SV (1997a) Alternative styles in the European wrasse, *Symphodus ocellatus*: Boldness-related schooling tendency. *Environ Biol Fish* 49:71-78
- Budaev SV (1997b) "Personality" in the guppy (*Poecilia reticulata*): A correlational study of exploratory behavior and social tendency. *J Comp Psychol* 111:399-411
- Budaev SV (2010) Using Principal Components and Factor Analysis in Animal Behaviour Research: Caveats and Guidelines. *Ethology* 116:472-480
- Burns JG (2008) The validity of three tests of temperament in guppies (*Poecilia reticulata*). *J Comp Psychol* 122:344-356
- Bushuev AV, Kerimov AB, Ivankina EV (2010) Estimation of heritability and repeatability of resting metabolic rate in birds, with free-living pied flycatchers *Ficedula hypoleuca* (Aves: Passeriformes) as an example. *Zh Obshch Biol* 71:402-424
- Carere C, Drent PJ, Privitera L, Koolhaas JM, Groothuis TGG (2005) Personalities in great tits, *Parus major*: stability and consistency. *Anim Behav* 70:795-805
- Carter AJ, Feeney WE, Marshall HH, Cowlshaw G, Heinsohn R (2013) Animal personality: what are behavioural ecologists measuring? *Biol Rev* 88:465-475
- Carvalho CF, Leitão AV, Funghi C, Batalha HR, Reis S, Mota PG, Lopes RJ, Cardoso GC (2013) Personality traits are related to ecology across a biological invasion. *Behav Ecol* 24:1081-1091
- Chervet N, Zöttl M, Schürch R, Taborsky M, Heg D (2011) Repeatability and Heritability of Behavioural Types in a Social Cichlid. *Int J Evol Biol* 2011: 321729
- Cote J, Clobert J, Brodin T, Fogarty S, Sih A (2010) Personality-dependent dispersal: characterization, ontogeny and consequences for spatially structured populations. *Philos T R Soc B* 365:4065-4076
- Dall SRX, Houston AI, McNamara JM (2004) The behavioural ecology of personality: consistent individual differences from an adaptive perspective. *Ecol Lett* 7:734-739
- David M, Auclair Y, Cezilly F (2012) Assessing short- and long-term repeatability and stability of personality in captive Zebra finches using longitudinal data. *Ethology* 118:932-942
- Dingemanse NJ, Both C, Drent PJ, Tinbergen JM (2004) Fitness consequences of avian personalities in a fluctuating environment. *Proc R Soc Lond B* 271:847-852
- Dingemanse NJ, Bouwman KM, van de Pol M, van Overveld T, Patrick SC, Matthysen E, Quinn JL (2012a) Variation in personality and behavioural plasticity across four populations of the great tit *Parus major*. *J Anim Ecol* 81:116-126
- Dingemanse NJ, Dochtermann NA, Nakagawa S (2012b) Defining behavioural syndromes and the role of 'syndrome deviation' in understanding their evolution. *Behav Ecol Sociobiol* 66:1543-1548
- Dingemanse NJ, Dochtermann NA, Wright J (2010) A method for exploring the structure of behavioural syndromes to allow formal comparison within and between data sets. *Anim Behav* 79:439-450
- Dingemanse NJ, Wright J, Kazem AJN, Thomas DK, Hickling R, Dawnay N (2007) Behavioural syndromes differ predictably between 12 populations of three-spined stickleback. *J Anim Ecol* 76:1128-1138
- Dochtermann NA, Jenkins SH (2007) Behavioural syndromes in Merriam's kangaroo rats (*Dipodomys merriami*): a test of competing hypotheses. *Proc R Soc Lond B* 274:2343-2349
- Dochtermann NA, Roff DA (2010) Applying a quantitative genetics framework to behavioural syndrome research. *Proc R Soc Lond B* 365:4013-4020
- Gilmour AR, Gogel BJ, Cullis BR, Thompson R (2009) ASReml user guide release 3.0. VSNi, Hemel Hempstead, UK
- Gosling SD (2001) From mice to men: What can we learn about personality from animal research? *Psychol Bull* 127:45-86

- Hadfield J (2010) MCMCglmm course notes. <http://cranr-project.org/web/packages/MCMCglmm/vignettes/CourseNotespdf>
- Hall CS (1934) Emotional behavior in the rat I Defecation and urination as measures of individual differences in emotionality. *J Comp Psychol* 18:385-403
- Huntingford FA (1976) Relationship between anti-predator behavior and aggression among conspecifics in 3-spined stickleback, *Gasterosteus aculeatus*. *Anim Behav* 24:245-260
- Kanda LL, Louon L, Straley K (2012) Stability in activity and boldness across time and context in captive Siberian dwarf hamsters. *Ethology* 118:518-533
- Koolhaas JM, Korte SM, De Boer SF, Van Der Vegt BJ, Van Reenen CG, Hopster H, De Jong IC, Ruis MAW, Blokhuis HJ (1999) Coping styles in animals: current status in behavior and stress-physiology. *Neurosci Biobehav R* 23:925-935
- Kurvers RHJM, Nolet BA, Prins HHT, Ydenberg RC, van Oers K (2012) Boldness affects foraging decisions in barnacle geese: an experimental approach. *Behav Ecol* 23:1155-1161
- Moretz JA (2003) Aggression and RHP in the northern swordtail fish, *Xiphophorus cortezi*: The relationship between size and contest dynamics in male-male competition. *Ethology* 109:995-1008
- Morrissey MB, Parker DJ, Korsten P, Pemberton JM, Kruuk LEB, Wilson AJ (2012) The prediction of adaptive evolution: empirical application of the second theorem of selection and comparison to the breeder's equation. *Evolution* 66:2399-2410
- Mutzel A, Dingemanse NJ, Araya-Ajoy YG, Kempenaers B (2013) Parental provisioning behaviour plays a key role in linking personality with reproductive success. *Proc R Soc Lond B* 280
- Nakagawa S, Schielzeth H (2010) Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. *Biol Rev* 85:935-956
- Réale D, Reader SM, Sol D, McDougall PT, Dingemanse NJ (2007) Integrating animal temperament within ecology and evolution. *Biol Rev* 82:291-318
- Roberts BW, DelVecchio WF (2000) The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychol Bull* 126:3-25
- Ronning B, Moe B, Bech C (2005) Long-term repeatability makes basal metabolic rate a likely heritable trait in the zebra finch *Taeniopygia guttata*. *J Exp Biol* 208:4663-4669
- Rudin FS, Briffa M (2011) Is boldness a resource-holding potential trait? Fighting prowess and changes in startle response in the sea anemone, *Actinia equina*. *Proc R Soc Lond B* 279:1904-1910
- Sih A, Bell A, Johnson JC (2004) Behavioral syndromes: an ecological and evolutionary overview. *Trends Ecol Evol* 19:372-378
- Stamps J, Groothuis TGG (2010) The development of animal personality: relevance, concepts and perspectives. *Biol Rev* 85:301-325
- Toms CN, Echevarria DJ, Jouandot DJ (2010) A methodological review of personality-related studies in Fish: Focus on the shy-bold axis of behavior. *Int J Comp Psychol* 23:1-25
- Walsh RN, Cummins RA (1976) Open-field test - critical review. *Psychol Bull* 83:482-504
- Warren EW, Callaghan S (1975) Individual differences in response to and open field test by guppy *Poecilia reticulata* (Peters). *J Fish Biol* 7:105-113
- Wilson AJ, Grimmer A, Rosenthal GG (2013) Causes and consequences of contest outcome: aggressiveness, dominance and growth in the sheephead swordtail, *Xiphophorus birchmanni*. *Behav Ecol Sociobiol*
- Wilson AJ, Réale D, Clements MN, Morrissey MM, Postma E, Walling CA, Kruuk LEB, Nussey DH (2010) An ecologist's guide to the animal model. *J Anim Ecol* 79:13-26

Fig. 1

a)



b)

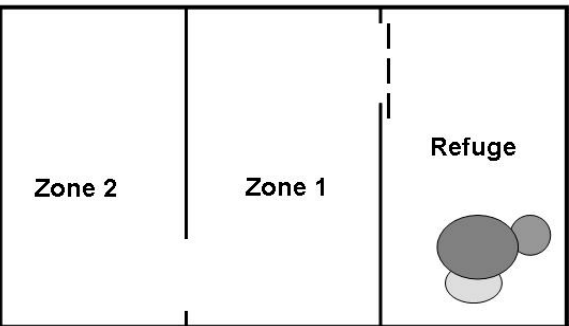
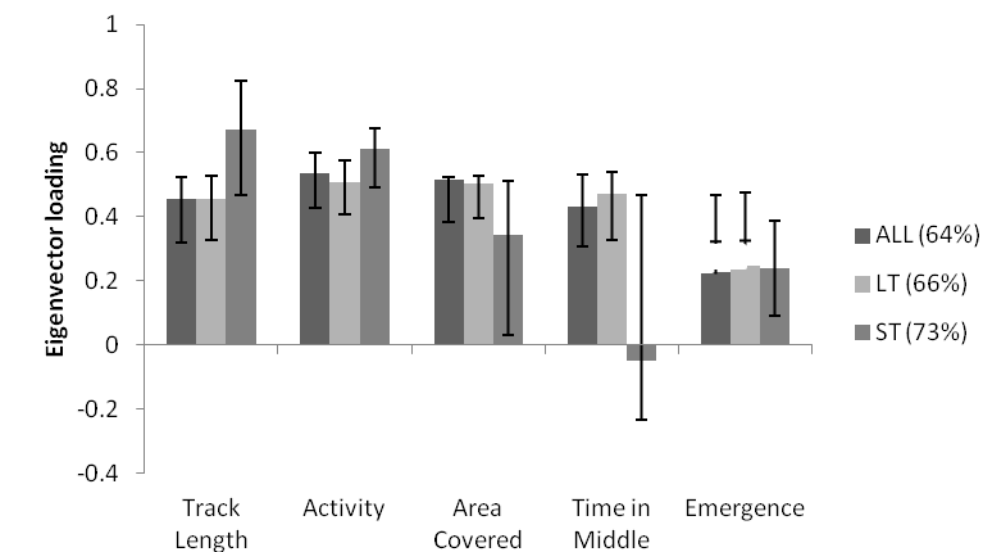


Fig. 2

a)



b)

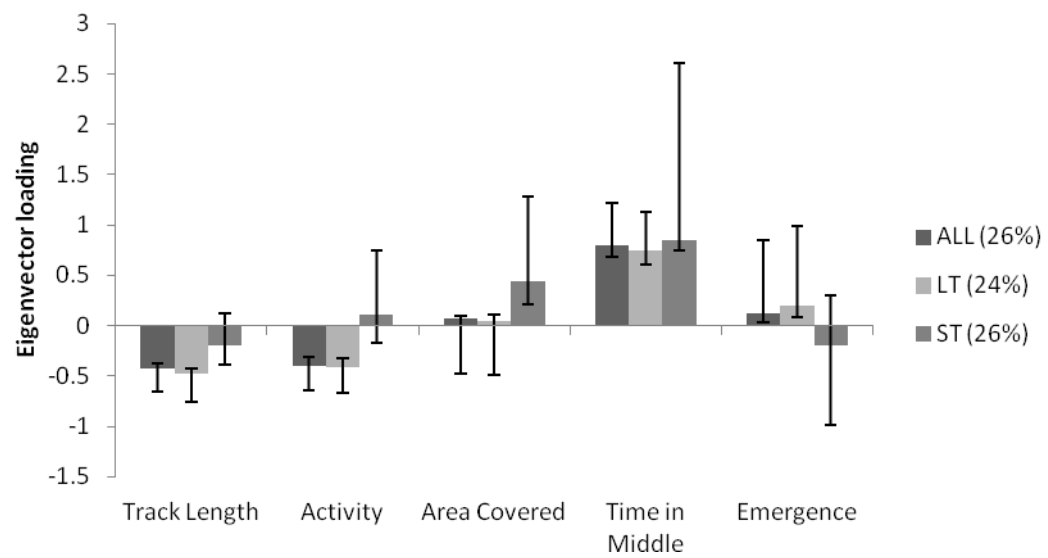


Fig. 3

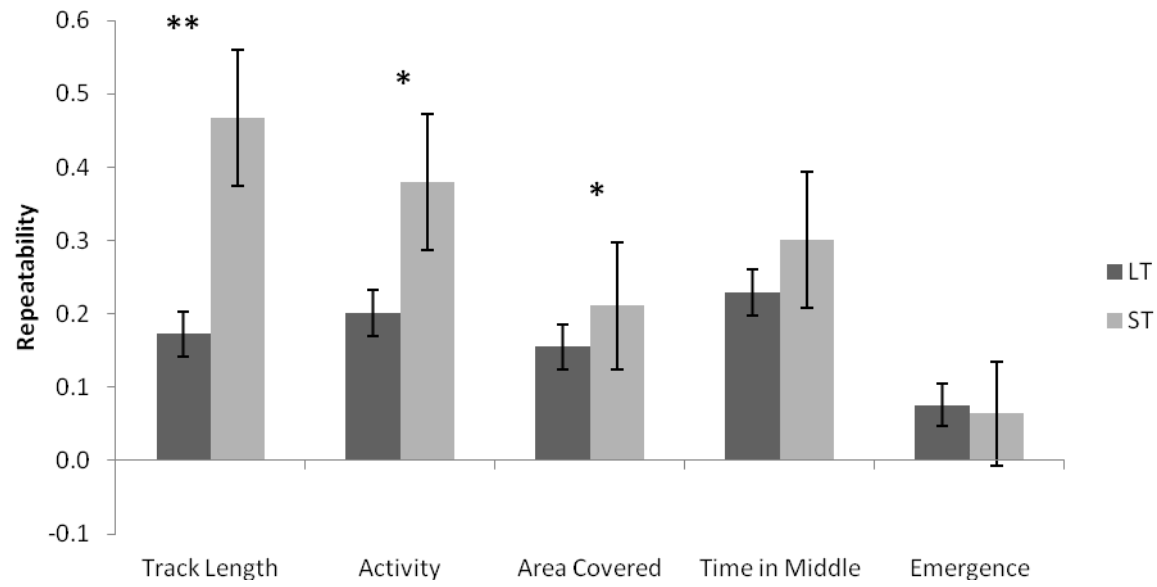
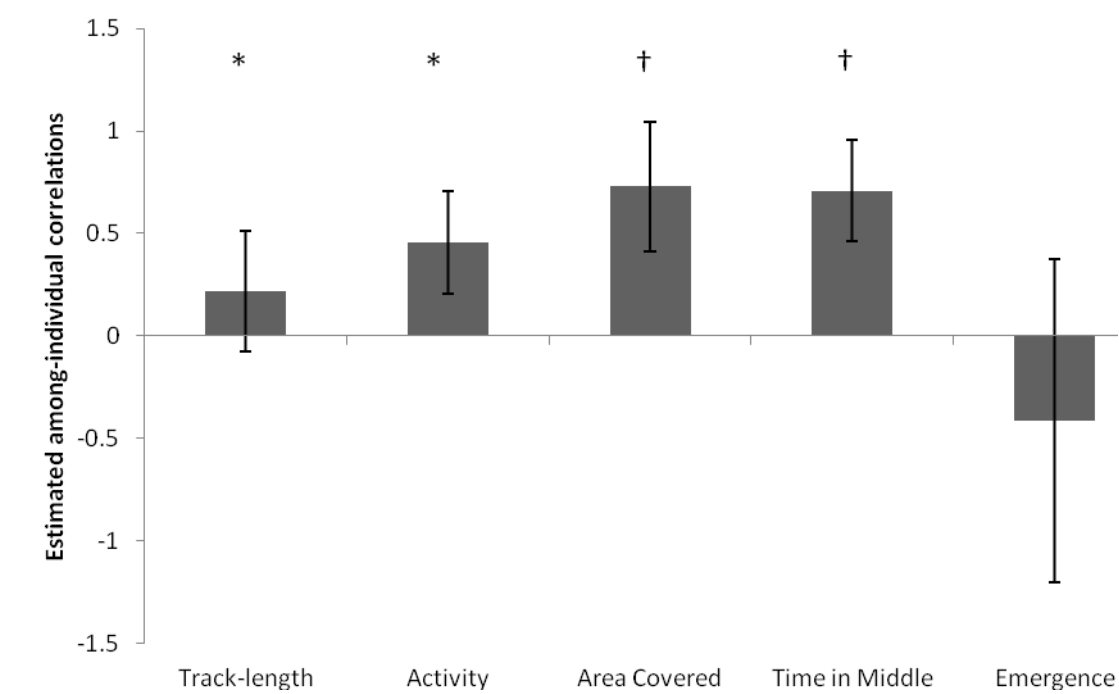


Fig. 4



Tables

Table 1 Data set for long-term (LT) and short-term (ST) studies. Number (N) and sex of individuals involved: male (M), female (F), total (T). Periods of data collection and intervals between trial pairs. Number of trials conducted: OFT (Open Field Trial); EET (Emergence & Exploration Trial); $N_{LT} = 2448$, $N_{ST} = 320$. Mean age of fish in days at the start of each trial pair with standard error in parentheses; “-“ indicates trial not performed

| Study | N | | | Data collection period | | | Number of trials | | Mean Fish Age (SE) | | | | |
|-------|-----|-----|-----|------------------------|----------|---------------------|------------------|------|--------------------|----------------|----------------|----------------|----------------|
| | M | F | T | Start | End | Days between trials | OFT | EET | 1 | 2 | 3 | 4 | 5 |
| LT | 223 | 150 | 373 | May 2011 | May 2012 | 56 | 1224 | 1224 | 203 (26.35) | 259 (26.44) | 372 (27.15) | 427 (27.13) | - |
| ST | 16 | 16 | 32 | Feb 2013 | Feb 2013 | 4 | 160 | 160 | 715 (13.36) | 719 (13.36) | 723 (13.36) | 727 (13.36) | 732 (13.36) |

Table 2 Behavioural traits recorded in OFT (Open Field Trials) and EET (Emergence and Exploration Trials)

| Trial type | Measured trait | Definition |
|------------|----------------------|--|
| OFT | Track Length (TL) | Distance swum (cm) |
| OFT | Activity (Act) | Percentage of time moving at a minimum 1.5cm/sec (%) |
| OFT | Area Covered (AC) | Area of tank floor covered (%) |
| OFT | Time in Middle (TIM) | Time spent in Zone 2 (seconds, see Figure 1) |
| EET | Emergence (Em) | Whether or not the fish emerged from the refuge (binary) |

Table 3 Among-individual variance/covariance matrix (**I**) from the multivariate analysis of a) all data, b) long-term study and c) short-term study. Estimates of variance (V_i , diagonal) with among-individual between-trait covariances (COV_i) below the diagonal and among-individual between-trait correlations (r_i ; above the diagonal). Standard errors are shown in parentheses for all parameter estimates. Traits: Track Length (TL), Activity (Act), Area Covered (AC), Time in Middle (TIM), Emergence (Em)

| a) All Data | TL | Act | AC | TIM | Em |
|---------------|----------------|---------------|---------------|----------------|----------------|
| TL | 0.130 (0.025) | 0.865 (0.033) | 0.750 (0.069) | 0.162 (0.117) | 0.304 (0.198) |
| Act | 0.124 (0.024) | 0.159 (0.026) | 0.731 (0.065) | 0.241 (0.106) | 0.577 (0.182) |
| AC | 0.097 (0.022) | 0.104 (0.022) | 0.128 (0.026) | 0.653 (0.075) | 0.414 (0.202) |
| TIM | 0.026 (0.019) | 0.042 (0.020) | 0.102 (0.023) | 0.192 (0.029) | 0.540 (0.180) |
| Em | 0.026 (0.017) | 0.054 (0.018) | 0.035 (0.018) | 0.056 (0.019) | 0.055 (0.024) |
| b) Long-term | TL | Act | Area | TIM | E |
| TL | 0.143 (0.028) | 0.892 (0.030) | 0.777 (0.069) | 0.238 (0.118) | 0.272 (0.192) |
| Act | 0.137 (0.026) | 0.164 (0.028) | 0.708 (0.072) | 0.314 (0.106) | 0.539 (0.180) |
| AC | 0.108 (0.025) | 0.106 (0.025) | 0.136 (0.030) | 0.704 (0.075) | 0.458 (0.208) |
| TIM | 0.041 (0.022) | 0.058 (0.022) | 0.118 (0.026) | 0.207 (0.033) | 0.607 (0.181) |
| Em | 0.027 (0.020) | 0.058 (0.020) | 0.045 (0.021) | 0.073 (0.022) | 0.071 (0.028) |
| c) Short-term | TL | Act | Area | TIM | E |
| TL | 0.458 (0.155) | 0.926 (0.041) | 0.640 (0.182) | -0.247 (0.256) | 1.070 (0.513) |
| Act | 0.381 (0.137) | 0.369 (0.134) | 0.812 (0.112) | 0.017 (0.274) | 1.001 (0.502) |
| AC | 0.188 (0.095) | 0.214 (0.097) | 0.188 (0.089) | 0.492 (0.222) | 0.545 (0.524) |
| TIM | -0.083 (0.089) | 0.005 (0.084) | 0.106 (0.079) | 0.248 (0.101) | -0.667 (0.557) |
| Em | 0.165 (0.080) | 0.139 (0.073) | 0.054 (0.056) | -0.076 (0.059) | 0.052 (0.066) |

How stable are personalities? A multivariate view of behavioural variation over long and short timescales in the sheephead swordtail, *Xiphophorus birchmanni*.

Behavioral Ecology and Sociobiology

Boulton K, Grimmer AJ, Rosenthal GG, Walling CA, Wilson AJ

Corresponding author: AJ Wilson, University of Exeter, A.Wilson@exeter.ed.ac

Supplemental Tables

Supplemental Table S1 MCMCglmm analyses of the binary Emergence trait. Table shows a) the intraclass correlation (IC - the binary equivalent of the repeatability (see methods)) from a univariate model, and b) among-individual correlation (r_I) estimates from bivariate models of Emergence and each open field trial trait

| Model | Trait(s) | | IC | r_I | 95% HPD interval | |
|-------|-----------|----------------|-------|-------|------------------|-------|
| | | | | | Lower | Upper |
| a) | Emergence | - | 0.090 | | 0.024 | 0.177 |
| b) | Emergence | Track Length | | 0.641 | 0.303 | 0.999 |
| | Emergence | Activity | | 0.736 | 0.488 | 0.977 |
| | Emergence | Area Covered | | 0.560 | 0.308 | 0.920 |
| | Emergence | Time in Middle | | 0.573 | 0.300 | 0.872 |

Supplemental Table S2 Univariate analyses of observed behavioural traits using the full (ALL), long- (LT) and short-term (ST) study data fitted using ASReml. The among- (V_I) and within-individual (residual) variance (V_R) estimates are presented for each trait along with repeatability (R). χ^2_1 and P-values relate to likelihood ratio tests of the significance of V_I . Note that for univariate models only we assume the test statistic to be asymptotically distributed as a mix of 50:50 χ^2_0 and χ^2_1 (following Visscher 2006). Behavioural traits studied: Track- length (TL), Activity (Act), Area Covered (AC), Time in Middle of tank (TIM), Emergence (Em). Behavioural traits studied: Track Length (TL), Activity (Act), Area Covered (AC), Time in Middle of tank (TIM), Emergence (Em)

| Data | Trait | V_I (SE) | V_R (SE) | R (SE) | χ^2_1 | P |
|------|-------|---------------|---------------|---------------|------------|--------|
| ALL | TL | 0.132 (0.025) | 0.658 (0.029) | 0.167 (0.029) | 56.10 | <0.001 |
| | Act | 0.159 (0.027) | 0.668 (0.029) | 0.193 (0.029) | 75.01 | <0.001 |
| | AC | 0.124 (0.026) | 0.767 (0.033) | 0.140 (0.027) | 41.86 | <0.001 |
| | TIM | 0.185 (0.029) | 0.682 (0.030) | 0.214 (0.029) | 82.61 | <0.001 |
| | Em | 0.058 (0.024) | 0.889 (0.039) | 0.061 (0.025) | 6.88 | 0.005 |
| LT | TL | 0.143 (0.028) | 0.689 (0.033) | 0.172 (0.031) | 41.51 | <0.001 |
| | Act | 0.165 (0.028) | 0.655 (0.031) | 0.201 (0.031) | 64.01 | <0.001 |
| | AC | 0.141 (0.030) | 0.768 (0.037) | 0.155 (0.031) | 31.76 | <0.001 |
| | TIM | 0.206 (0.033) | 0.693 (0.033) | 0.229 (0.032) | 69.85 | <0.001 |
| | Em | 0.072 (0.028) | 0.887 (0.043) | 0.075 (0.029) | 7.87 | 0.003 |
| ST | TL | 0.457 (0.154) | 0.520 (0.067) | 0.468 (0.093) | 41.28 | <0.001 |
| | Act | 0.369 (0.133) | 0.571 (0.073) | 0.393 (0.095) | 29.23 | <0.001 |
| | AC | 0.186 (0.089) | 0.663 (0.085) | 0.220 (0.089) | 10.08 | 0.002 |
| | TIM | 0.248 (0.101) | 0.594 (0.076) | 0.295 (0.093) | 17.20 | <0.001 |
| | Em | 0.061 (0.069) | 0.885 (0.113) | 0.064 (0.071) | 1.03 | 0.156 |

Supplemental Table S3 Estimates of fixed effects (with standard errors in parentheses) from univariate mixed models of each behavioural trait for the data combined and for the long- (LT) and short-term (ST) studies. Significance was assessed using conditional F statistics and all models contained a random effect of individual identity. Coefficients are not presented for Stack, Treatment and Trial*Treatment due to their being multilevel factors. All individuals from ST were housed in the same stack therefore this covariate was not included in the ST analyses. Traits: Track Length (TL), Activity (Act), Area Covered (AC), Time in Middle (TIM), Emergence (Em).

| Dataset | Response | Fixed Effect | Coefficient (SE) | DF | F | P |
|---------|----------|-----------------|------------------|----------|---------|--------|
| All | TL | Mean | 1.898 (0.121) | 1,332.7 | 1666.86 | <0.001 |
| | | Sex | -0.046 (0.061) | 1,348.8 | 0.56 | 0.454 |
| | | Day order | -0.006 (0.002) | 1,1376.1 | 8.01 | 0.005 |
| | | Stack | | 6,545.5 | 53.11 | <0.001 |
| | | Trial | 0.230 (0.028) | 1,1126.2 | 207.67 | <0.001 |
| | | Treatment | | 3,339.3 | 1.56 | 0.201 |
| | | Trial*Treatment | | 3,1375.9 | 2.68 | 0.046 |
| | Act | Mean | 3.223 (0.125) | 1,347.1 | 3860.25 | <0.001 |
| | | Sex | -0.145 (0.064) | 1,365.2 | 5.07 | 0.026 |
| | | Day order | -0.003 (0.002) | 1,1373 | 1.96 | 0.164 |
| | | Stack | | 6,564.6 | 33.46 | <0.001 |
| | | Trial | 0.238 (0.028) | 1,1129.5 | 225.4 | <0.001 |
| | | Treatment | | 3,353.4 | 3.86 | 0.01 |
| | | Trial*Treatment | | 3,1374.5 | 4.75 | 0.003 |
| | AC | Mean | 2.795 (0.127) | 1,339.8 | 2204.17 | <.001 |
| | | Sex | 0.252 (0.063) | 1,354.5 | 15.76 | <0.001 |
| | | Day order | -0.006 (0.002) | 1,1376.9 | 7.57 | 0.006 |
| | | Stack | | 6,555.8 | 7.87 | <0.001 |
| | | Trial | 0.179 (0.030) | 1,1141.5 | 112.28 | <0.001 |
| | | Treatment | | 3,347.2 | 2.37 | 0.071 |
| | | Trial*Treatment | | 3,1363.8 | 2.24 | 0.083 |
| | TIM | Mean | 1.579 (0.128) | 1,342.4 | 621.85 | <0.001 |
| | | Sex | 0.528 (0.067) | 1,361.7 | 62.39 | <0.001 |

| | | | | | | | |
|-----------------|-----------|-----------------|----------------|----------------|---------|--------|--------|
| | | Day order | -0.009 (0.002) | 1,1368.7 | 15.56 | <0.001 | |
| | | Stack | | 6,559.8 | 9.52 | <0.001 | |
| | | Trial | 0.075 (0.029) | 1,1119.7 | 10.36 | 0.001 | |
| | | Treatment | | 3,348.2 | 0.85 | 0.47 | |
| | | Trial*Treatment | | 3,1367.3 | 6.13 | <0.001 | |
| | Em | Mean | 0.665 (0.130) | 1,297.7 | 141.35 | <0.001 | |
| | | Sex | 0.222 (0.060) | 1,301.2 | 13.54 | <0.001 | |
| | | Day order | 0.007 (0.003) | 1,1342 | 6.21 | 0.007 | |
| | | Stack | | 6,525.1 | 9.53 | <0.001 | |
| | | Trial | -0.085 (0.032) | 1,1138 | 8.31 | 0.004 | |
| | | Treatment | | 3,306.7 | 2.63 | 0.051 | |
| | | Trial*Treatment | | 3,1111.1 | 1.07 | 0.048 | |
| | LT | TL | Mean | 1.723 (0.144) | 1,350.1 | 611.49 | <0.001 |
| | | | Sex | -0.043 (0.065) | 1,348 | 0.44 | 0.505 |
| Day order | | | -0.007 (0.002) | 1,1219.6 | 9.8 | 0.002 | |
| Stack | | | | 5,353.6 | 7.81 | <0.001 | |
| Trial | | | 0.310 (0.043) | 1,976.1 | 226.07 | <0.001 | |
| Treatment | | | | 3,354.8 | 1.34 | 0.263 | |
| Trial*Treatment | | | | 3,980.2 | 0.26 | 0.853 | |
| Act | | Mean | 3.065 (0.143) | 1,350.7 | 2106.73 | <0.001 | |
| | | Sex | -0.164 (0.066) | 1,349.2 | 6.17 | 0.014 | |
| | | Day order | -0.004 (0.002) | 1,1219.5 | 3.16 | 0.078 | |
| | | Stack | | 5,354.1 | 9.25 | <0.001 | |
| | | Trial | 0.311 (0.042) | 1,969.7 | 242.22 | <0.001 | |
| | | Treatment | | 3,355.2 | 2.89 | 0.036 | |
| | | Trial*Treatment | | 3,973.6 | 0.67 | 0.571 | |
| AC | Mean | 2.707 (0.150) | 1,341.7 | 1466.43 | <0.001 | | |
| | Sex | 0.244 (0.067) | 1,339.4 | 13.26 | <0.001 | | |
| | Day order | -0.007 (0.002) | 1,1217.9 | 8.15 | 0.005 | | |
| | Stack | | 5,345.4 | 5.21 | <0.001 | | |
| | Trial | 0.282 (0.045) | 1,973.9 | 99.55 | <.001 | | |

| | | | | | | |
|-----------------|-----|-----------------|----------------|----------|--------|--------|
| | | Treatment | | 3,346.5 | 1.32 | 0.27 |
| | | Trial*Treatment | | 3,978.1 | 1.42 | 0.237 |
| | TIM | Mean | 1.667 (0.150) | 1,349.4 | 588.12 | <0.001 |
| | | Sex | 0.540 (0.071) | 1,348.4 | 58.39 | <0.001 |
| | | Day order | -0.010 (0.002) | 1,1216.3 | 17.09 | <0.001 |
| | | Stack | | 5,352.6 | 8.49 | <0.001 |
| | | Trial | 0.075 (0.043) | 1,962.3 | 2.3 | 0.132 |
| | | Treatment | | 3,353.7 | 1.27 | 0.285 |
| | | Trial*Treatment | | 3,966.2 | 3.03 | 0.029 |
| | Em | Mean | 0.654 (0.155) | 1,336.3 | 143.76 | <0.001 |
| | | Sex | 0.198 (0.064) | 1,330.8 | 9.46 | 0.002 |
| | | Day order | 0.009 (0.003) | 1,1179.3 | 7.51 | 0.007 |
| | | Stack | | 5,342.4 | 4.67 | <0.001 |
| | | Trial | -0.085 (0.049) | 1,983.7 | 7.9 | 0.005 |
| | | Treatment | | 3,340.8 | 1.77 | 0.153 |
| Trial*Treatment | | | 3,995.8 | 1.39 | 0.244 | |
| ST | TL | Mean | 2.487 (0.508) | 1,26.9 | 33.21 | <0.001 |
| | | Sex | 0.064 (0.267) | 1,38.2 | 0.06 | 0.81 |
| | | Day order | 0.013 (0.007) | 1,122.2 | 4.11 | 0.046 |
| | | Trial | -0.029 (0.065) | 1,121 | 5.12 | 0.027 |
| | | Treatment | | 3,27.2 | 0.72 | 0.547 |
| | | Trial*Treatment | | 3,120.9 | 4.35 | 0.006 |
| | Act | Mean | 3.144 (0.521) | 1,26.9 | 69.66 | <0.001 |
| | | Sex | 0.270 (0.253) | 1,35.7 | 1.14 | 0.292 |
| | | Day order | 0.011 (0.007) | 1,122.6 | 2.78 | 0.1 |
| | | Trial | 0.031 (0.068) | 1,121.1 | 7.58 | 0.007 |
| | | Treatment | | 3,27.1 | 0.99 | 0.411 |
| | | Trial*Treatment | | 3,121 | 3.36 | 0.021 |
| | AC | Mean | 1.263 (0.542) | 1,27 | 17.17 | <0.001 |
| | | Sex | 0.394 (0.212) | 1,32.1 | 3.45 | 0.072 |
| | | Day order | 0.005 (0.007) | 1,124.4 | 0.51 | 0.474 |

| | | | | | | |
|--|-----|-----------------|----------------|---------|-------|--------|
| | | Trial | 0.145 (0.073) | 1,121.7 | 14.43 | <0.001 |
| | | Treatment | | 3,27.1 | 2.82 | 0.058 |
| | | Trial*Treatment | | 3,121.5 | 0.65 | 0.588 |
| | TIM | Mean | -0.027 (0.521) | 1,27.1 | 3.99 | 0.056 |
| | | Sex | 0.447 (0.224) | 1,33.5 | 3.98 | 0.054 |
| | | Day order | 0.001 (0.007) | 1,123.6 | 0.01 | 0.904 |
| | | Trial | 0.246 (0.069) | 1,121.5 | 17.04 | <0.001 |
| | | Treatment | | 3,27.2 | 2.87 | 0.056 |
| | | Trial*Treatment | | 3,121.4 | 1.46 | 0.231 |
| | Em | Mean | 1.534 (0.612) | 1,26.9 | 19.29 | <0.001 |
| | | Sex | 0.438 (0.186) | 1,28 | 5.53 | 0.026 |
| | | Day order | -0.005 (0.008) | 1,135.4 | 0.43 | 0.512 |
| | | Trial | -0.062 (0.083) | 1,125.7 | 0.73 | 0.395 |
| | | Treatment | | 3,27 | 1.69 | 0.193 |
| | | Trial*Treatment | | 3,123.1 | 0.79 | 0.502 |

Supplemental Table S4 Tests for heterogeneity of variance structures across density treatments for each behavioural trait in the LT data sets. Presented are χ^2 statistics with associated P-values for comparing models with homogeneous and heterogeneous (i.e. treatment specific) among-individual and residual variances. Significant heterogeneity of variance components across density treatments is indicated for Track Length (TL) and Time in Middle (TIM) only. Treatment specific variance components for TL estimated under the heterogeneous model (not shown) demonstrate lower repeatability (SE) in the High/High treatment $R = 0.122$ (0.075), relative to other treatment classes (Low/Low $R = 0.311$ (0.073), Low/High $R = 0.209$ (0.087), High/Low $R = 0.273$ (0.097)). For TIM, repeatability (SE) is reduced in the Low/High treatment $R = 0.063$ (0.037), relative to other treatment classes (Low/Low $R = 0.243$ (0.050), High/High $R = 0.212$ (0.067), High/Low $R = 0.206$ (0.061))

| Trait | χ^2_6 | P |
|----------------|------------|--------|
| Track Length | 38.31 | <0.001 |
| Activity | 7.18 | 0.30 |
| Area Covered | 1.99 | 0.92 |
| Time in Middle | 20.96 | 0.002 |
| Emergence | 2.07 | 0.91 |

How stable are personalities? A multivariate view of behavioural variation over long and short timescales in the sheephead swordtail, *Xiphophorus birchmanni*.

Behavioral Ecology and Sociobiology

Boulton K, Grimmer AJ, Rosenthal GG, Walling CA, Wilson AJ

Corresponding author: AJ Wilson, University of Exeter, A.Wilson@exeter.ed.ac

Supplemental Figure Legends

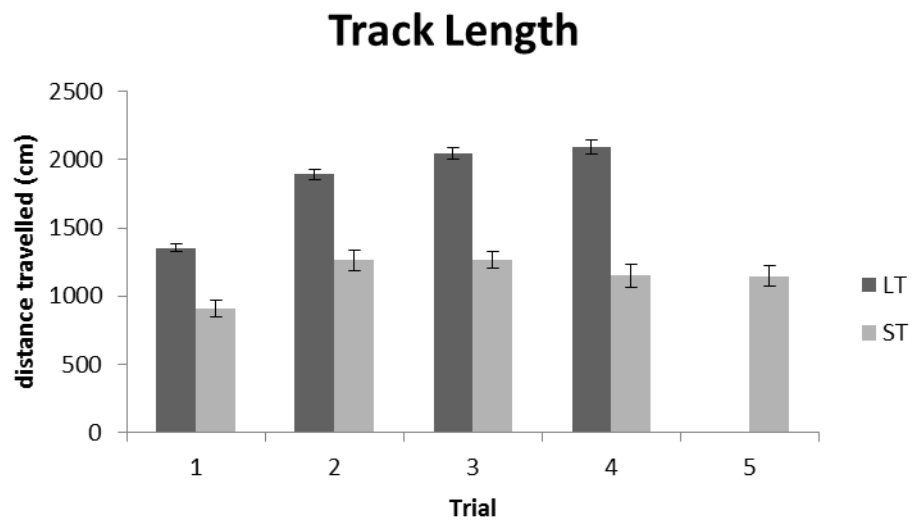
Fig. S1 Summary of raw behavioural data showing observed mean (\pm standard error) by Trial in long- (dark grey) and short- (light grey) term studies for a) Track Length, b) Activity, c) Area covered, d) Time in middle, and e) Emergence, which is represented as a percentage and therefore does not have an associated error. The long-term study (LT) comprised four Trials, while there were five Trials in the short-term (ST) study

Fig. S2. Posterior distribution of the intra-class correlation for the binary trait of Emergence modelled in MCMCglmm. The posterior mode for the intraclass correlation, IC = 0.109, 95% HPD interval 0.041 – 0.194

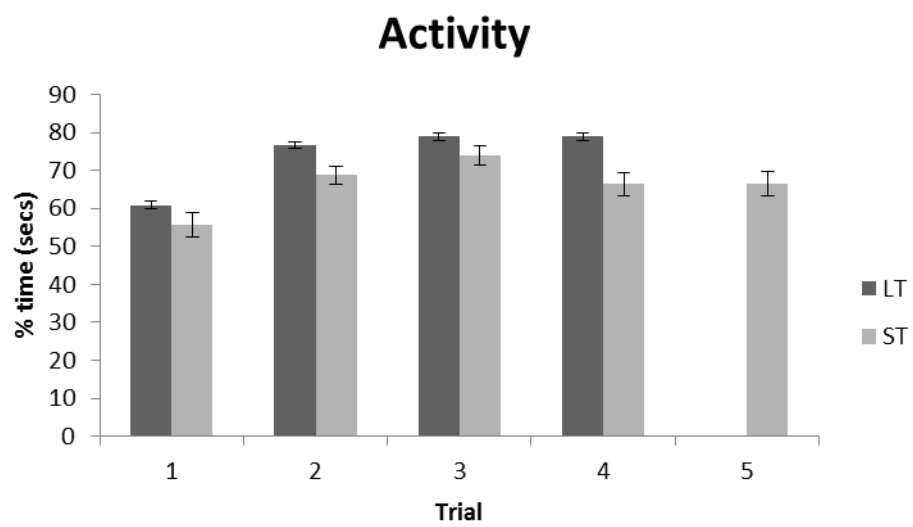
Fig. S3 Parametric bootstrap distributions for θ , the estimated angle between $EV1_{LT}$ and $EV1_{ST}$ in the case that a) I_{LT} and I_{ST} are equal to their REML estimates, and b) I_{LT} and I_{ST} are equal such that the true angle between leading eigenvectors is zero. Distributions are based on 5000 pairs of simulated matrices (see main text for further details)

Fig. S1

a)

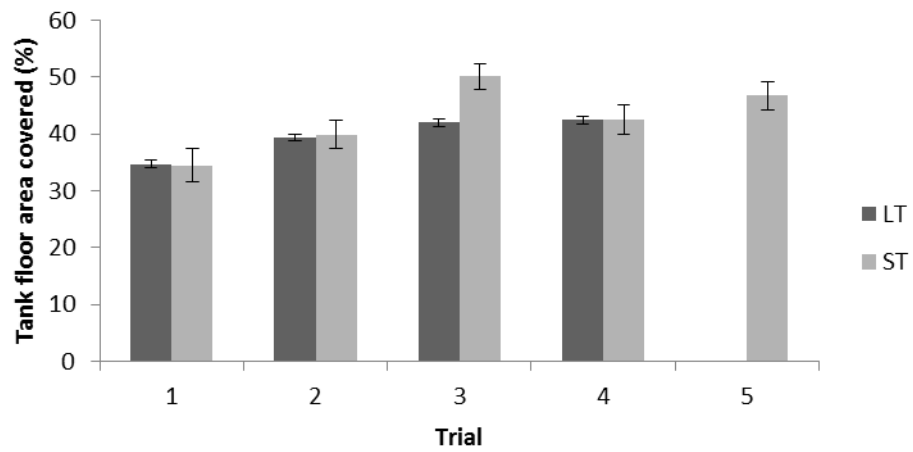


b)



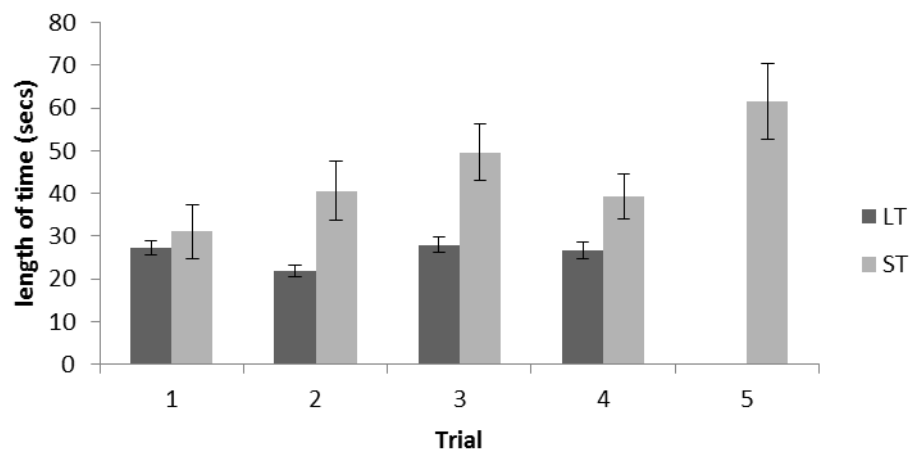
c)

Area Covered



d)

Time in Middle



e)

Emergence

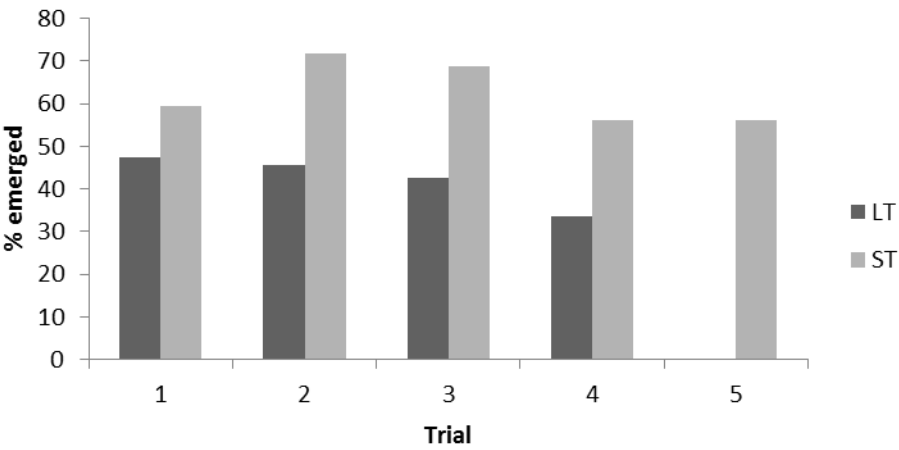


Fig. S2

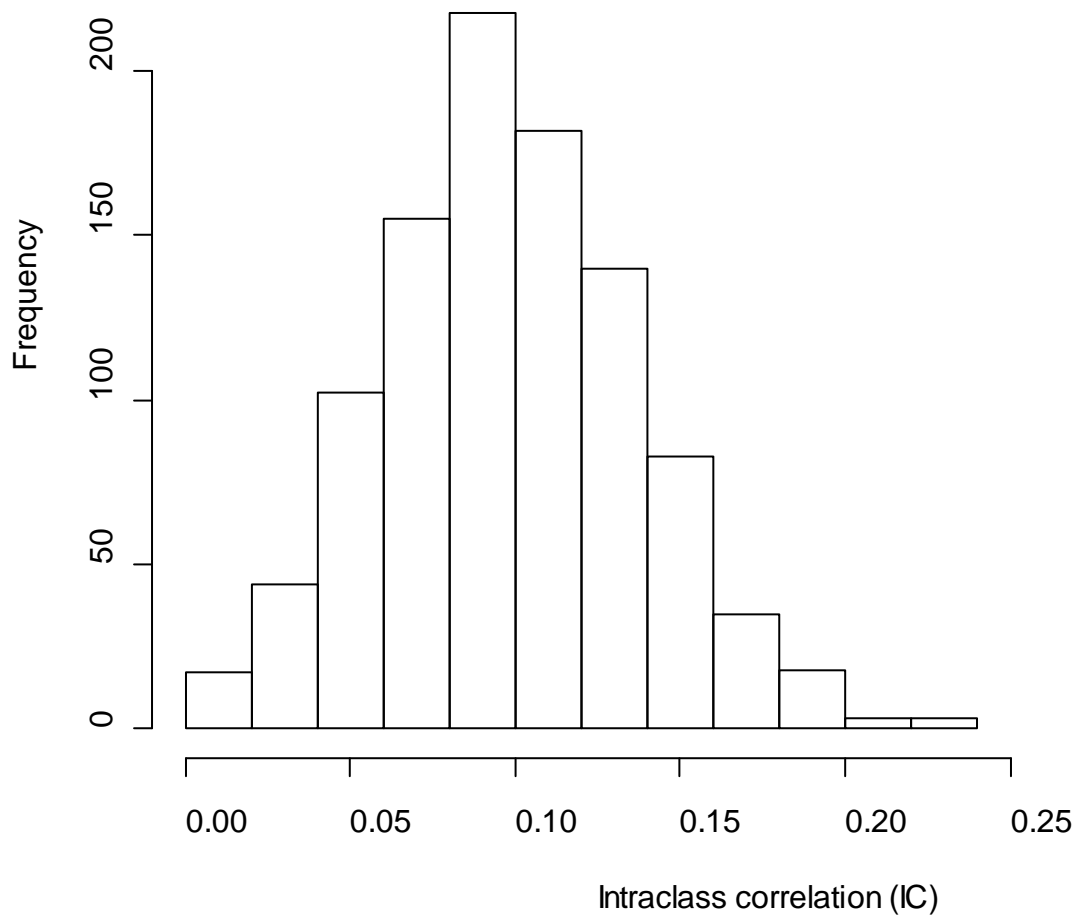
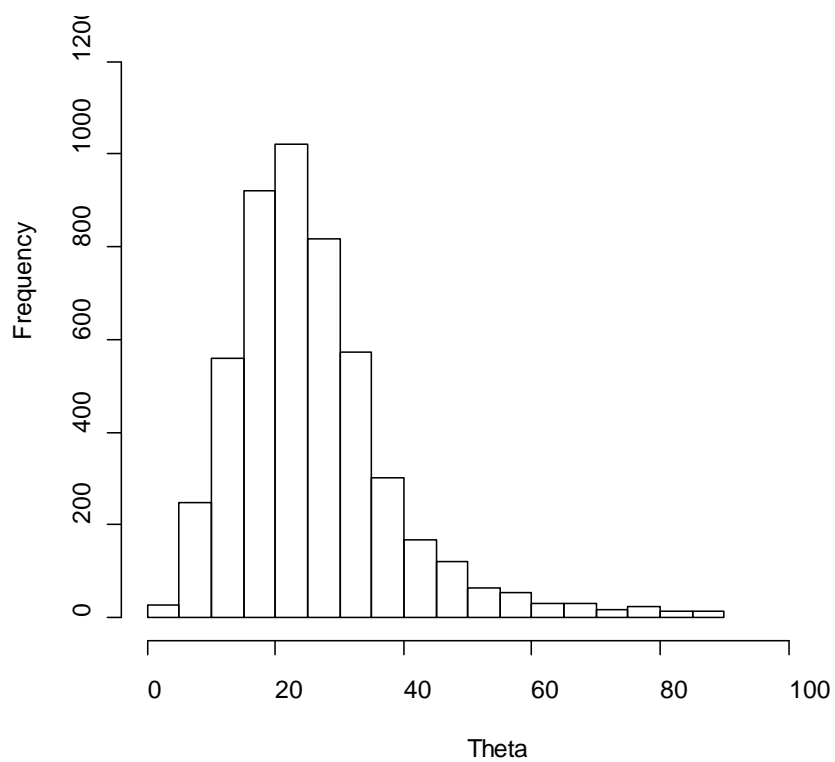


Fig. S3

a)



b)

